

# Phylogenomics:

## A Genome Level Approach to Assembling the Bacterial Branches of the Tree of Life

Jonathan A. Eisen

Naomi Ward

Karen E. Nelson

Frank T. Robb

Jonathan H. Badger

James Sakwa

Martin Wu

Dongying Wu

Kevin Penn

Elizabeth M. O'Connor

Julie Enticknap

Tim Steppe

<http://www.tigr.org/tol>

# Background I: rRNA Tree

- Phenotype not very useful for bacterial phylogeny
- Most molecular studies based on 16s rRNA sequence analysis
- Studies of other genes do not always agree with rRNA, especially for deep branches

## Background II: Most Bacteria Have Never Been Cultured

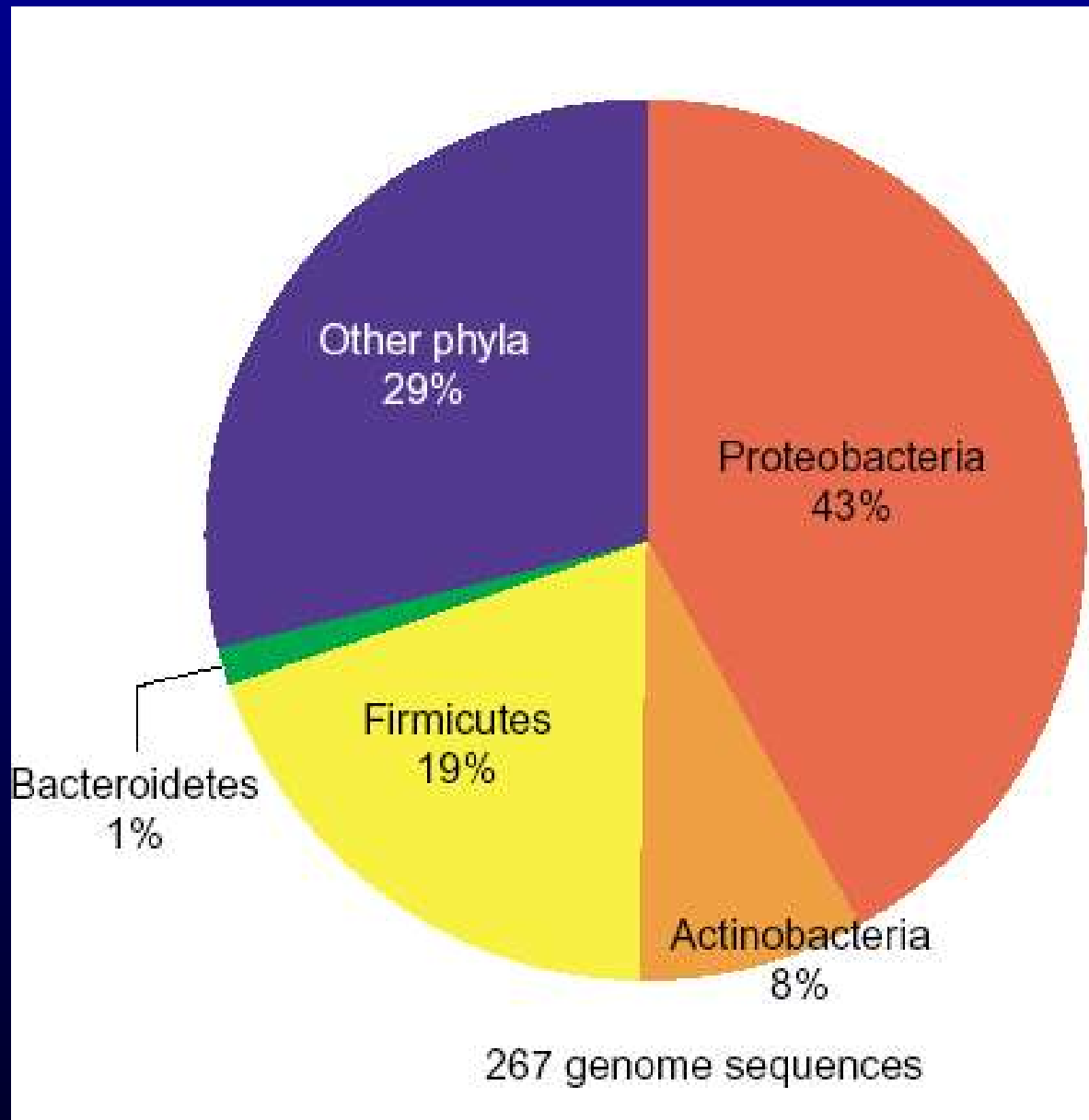
- Microscopic and molecular studies show that  $<1\%$  of the microbes in most environments have been grown in pure culture
- True in terms of #s and phylogenetic diversity
- This means we know little about their biology

## Background III:

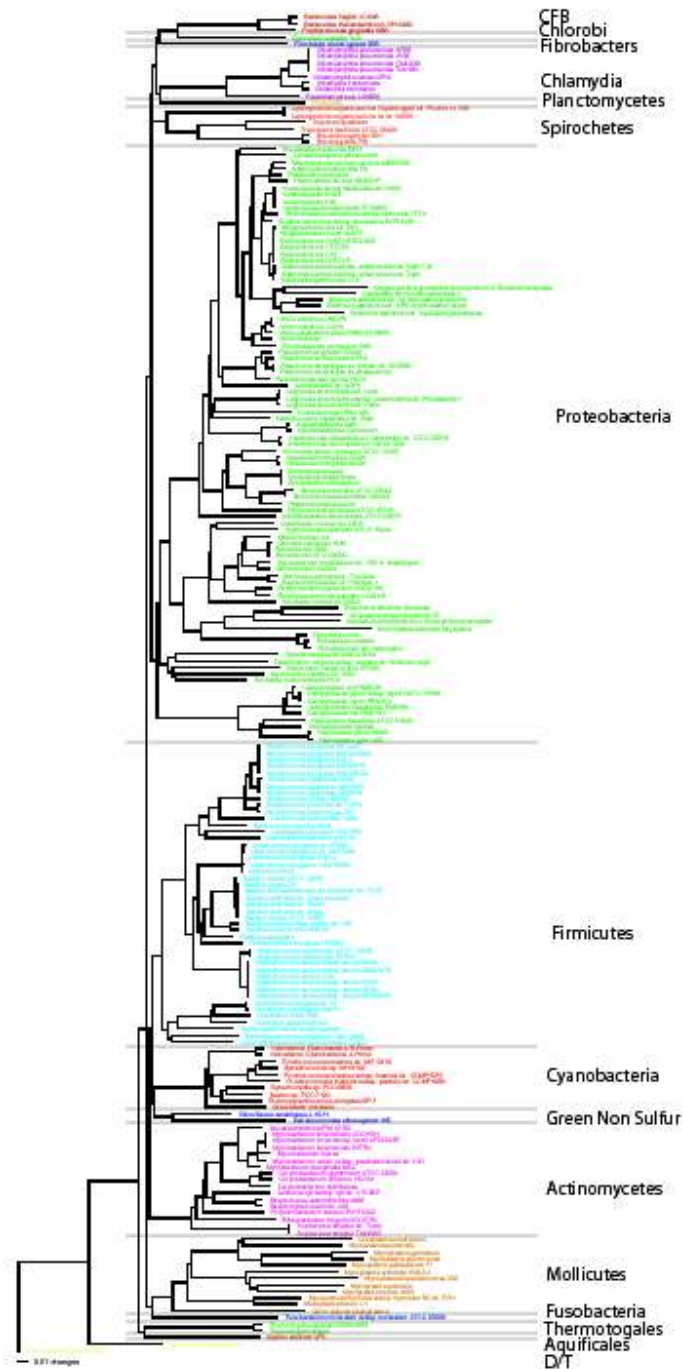
# Genomics Has Revolutionized Bacteriology

- Predictions of biology
- Drug design, vaccine development
- Functional genomic studies
- Evolutionary reconstructions
  - Whole genome phylogeny
  - Lateral gene transfer
  - Population genomics

# Biased Sampling of Bacterial Genomes



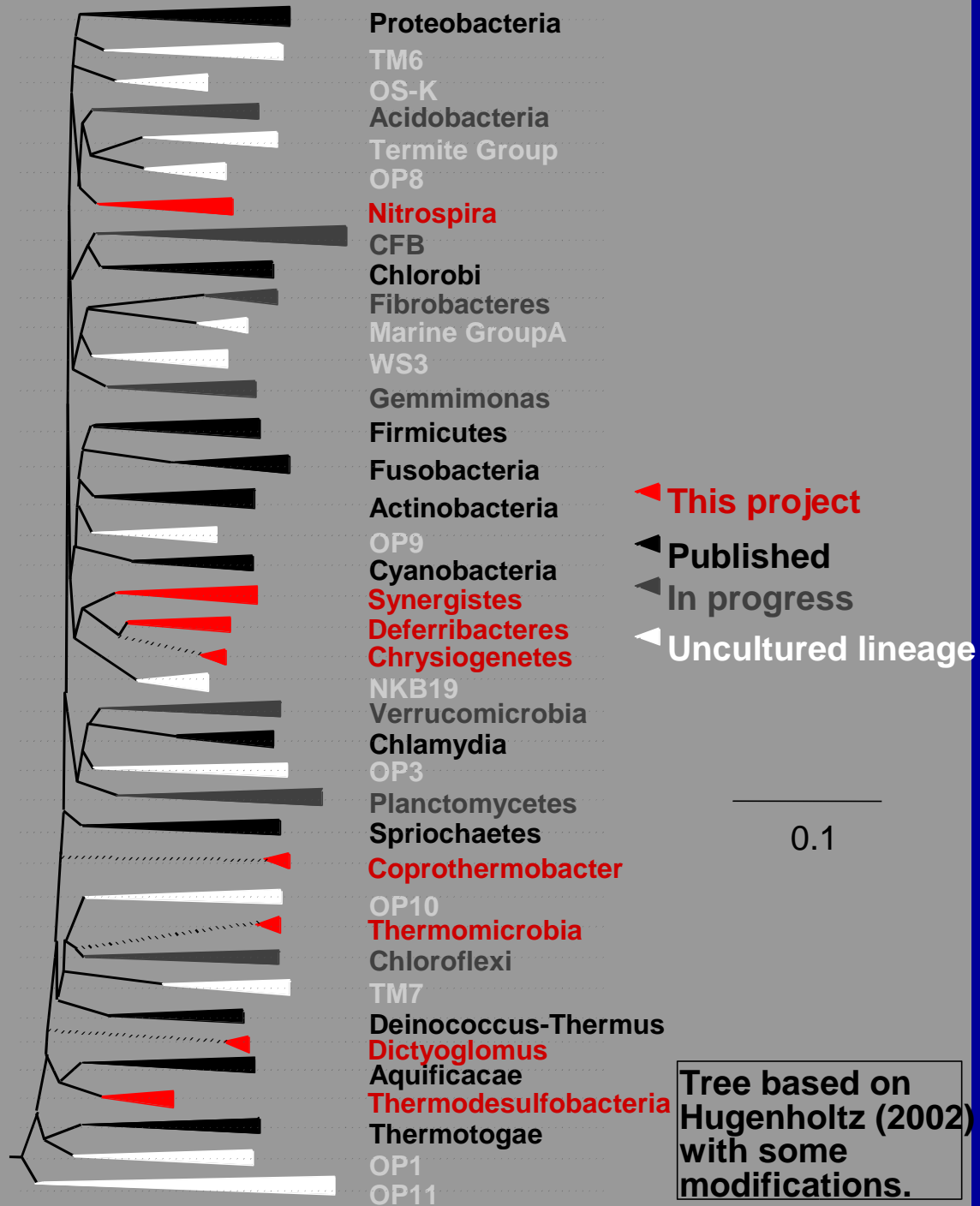
Hugenholtz  
2002



# Automated Whole Genome Phylogeny

# TIGR Tree of Life Project

- Major goal
  - Increase phylogenetic diversity of genome sequences
- Three sub-goals
  - Resolve relationships among the phyla
  - Launch experimental studies of these phyla
  - Inform environmental studies of uncultured microbes
- The Players
  - TIGR (Jonathan Eisen, Naomi Ward, Karen Nelson et al.)
  - COMB (Frank Robb et al.)



Tree based on Hugenholtz (2002) with some modifications.

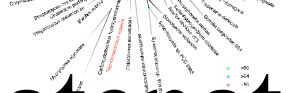


# Genome Sequencing Progress

<u>Phylum</u>	<u>Species selected</u>	<u>Growth, DNA isolation</u>	<u>Libraries</u>	<u>Shotgun Coverage</u>	<u>Estimated Genome Size(Mb)</u>	<u># of Contigs</u>	<u>Auto-Annotated</u>
Chrysiogenes	<i>Chrysiogenes arsenatis</i>	+	+	4x	2.5	155	+
Coprothermobacter	<i>Coprothermobacter proteolyticus</i> (CP)	+	+	8x	1.38	3	+
Dictyoglomi	<i>Dictyoglomusthermophilum</i> (DT)	+	+	8x	2.0	9	+
Thermodesulfobacteria	<i>Thermodesulfobacterium commune</i> (TC)	+	+	8x	1.78	26	+
Nitrospirae	<i>Thermodesulfovibrio yellowstonii</i> (TY)	+	+	8x	1.98	27	+
Thermomicrobia	<i>Thermomicrobium roseum</i>	+	+	8x	3.4	82	+
Deferribacteres	Selecting from <i>Deferribacter thermophilus</i> , <i>Geovibrio thiophilus</i> , <i>Flexistipes sinuatus</i>	+	In progress				
Synergistes	Selecting from <i>Synergistes jonesii</i> , <i>Aminobacter combiense</i> , <i>Thermanaerovibrio acidaminovorans</i> , <i>Aminomonas paucivorans</i> , <i>Dethiosulfovibrio peptidovorans</i>	+	In progress				

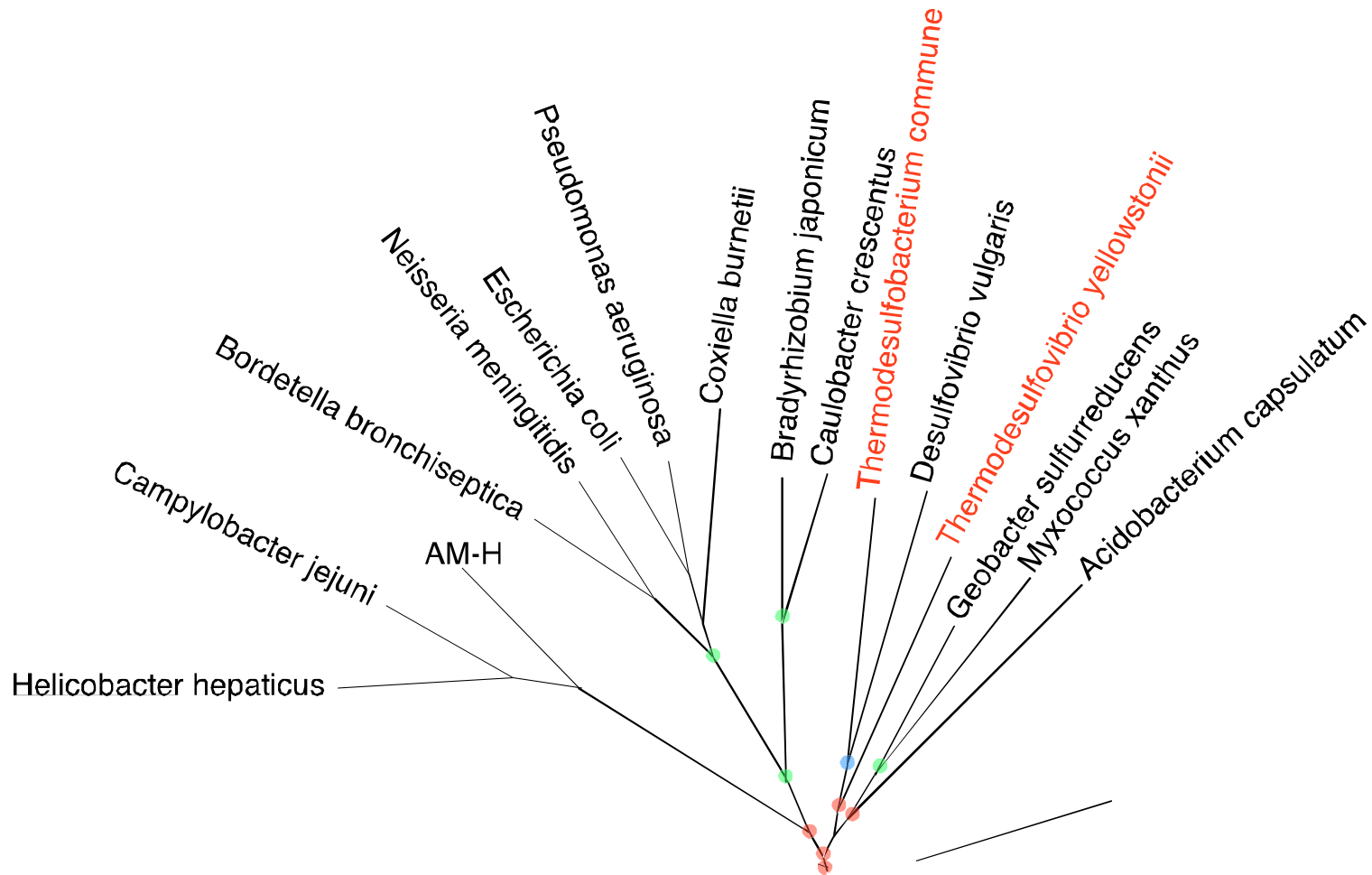
Goal I:

Relationships Among Phyla

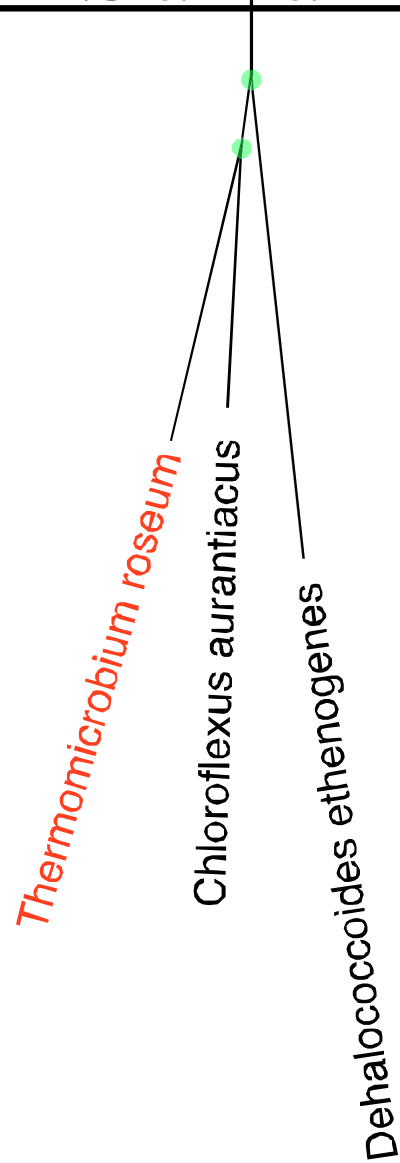


# Concatenated Alignment ML Tree

# Proteobacteria



# Green Non Sulfur Bacteria



Goal II:

Biology of These Phyla



Goal III:

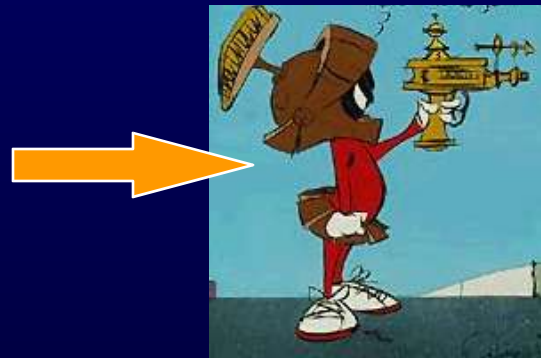
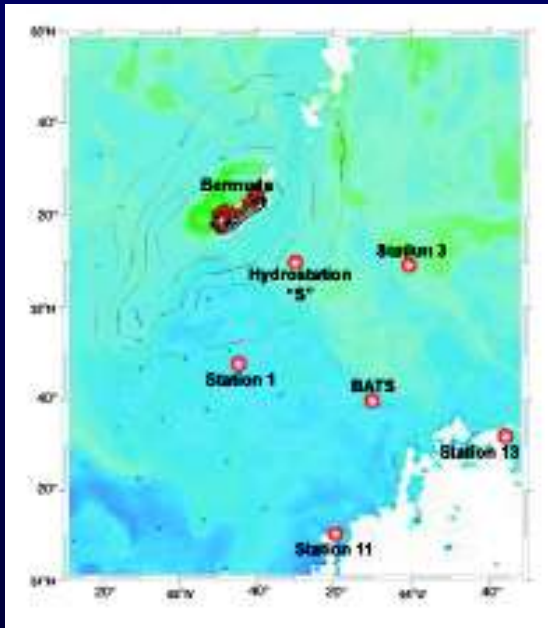
Uncultured Microbes



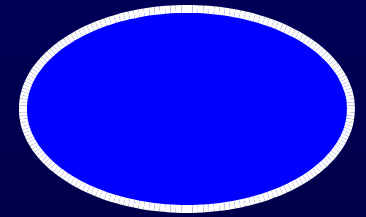
# Key Issues in Uncultured Microbes

- Questions
  - 1. Who is out there?
  - 2. What are they doing?
  - 3. Need to connect 1 and 2.
- Answers
  - 1. Phylogenetic anchors
  - 2. Genomics
  - 3. Linking anchors to genomics contigs

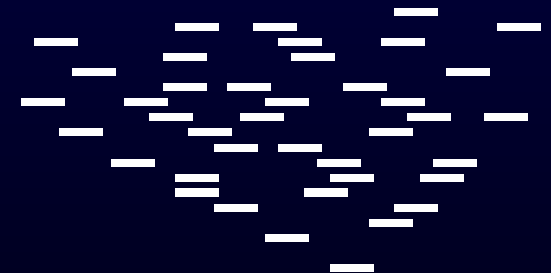
# Phylogenetic Anchors and the Sargasso Sea Shotgun Sequencing



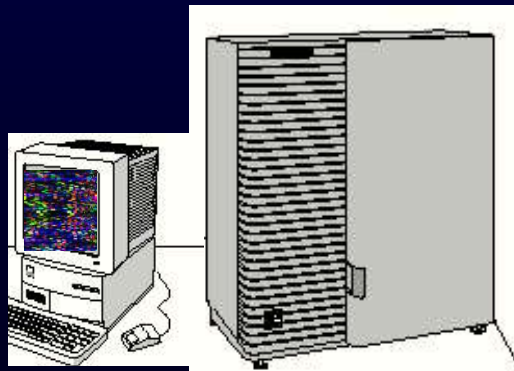
Warner Brothers, Inc.



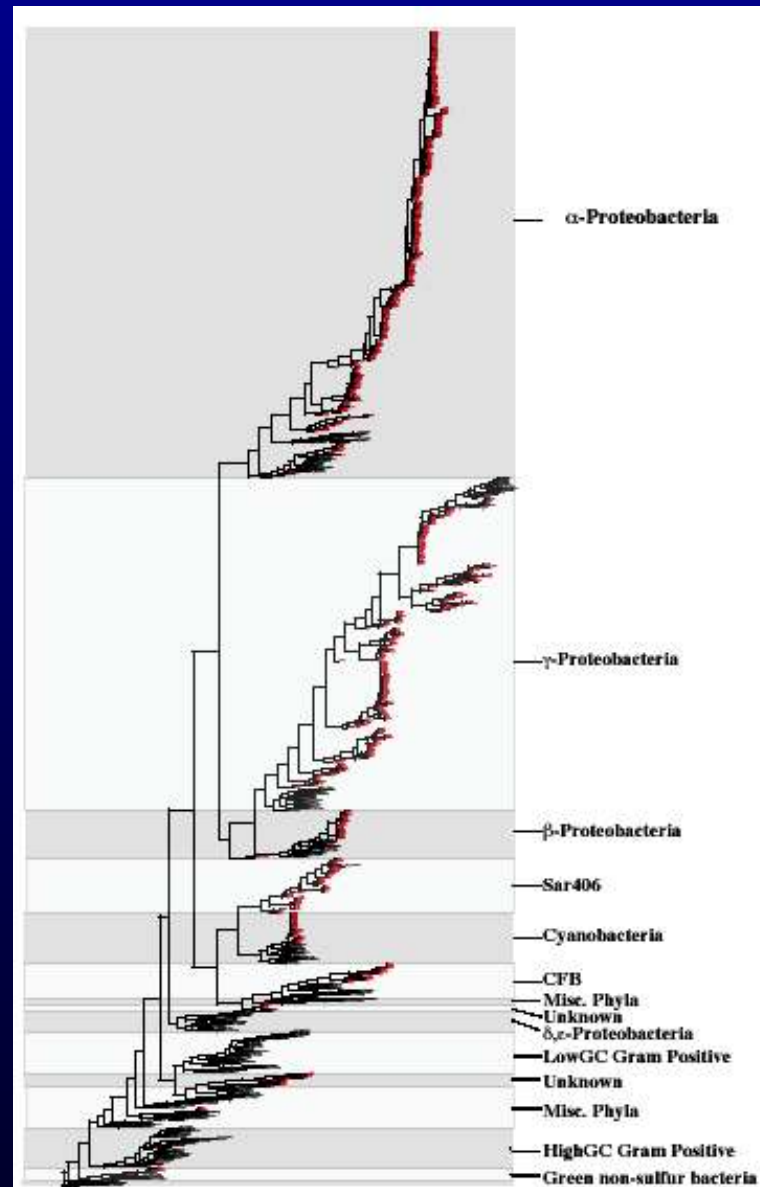
shotgun



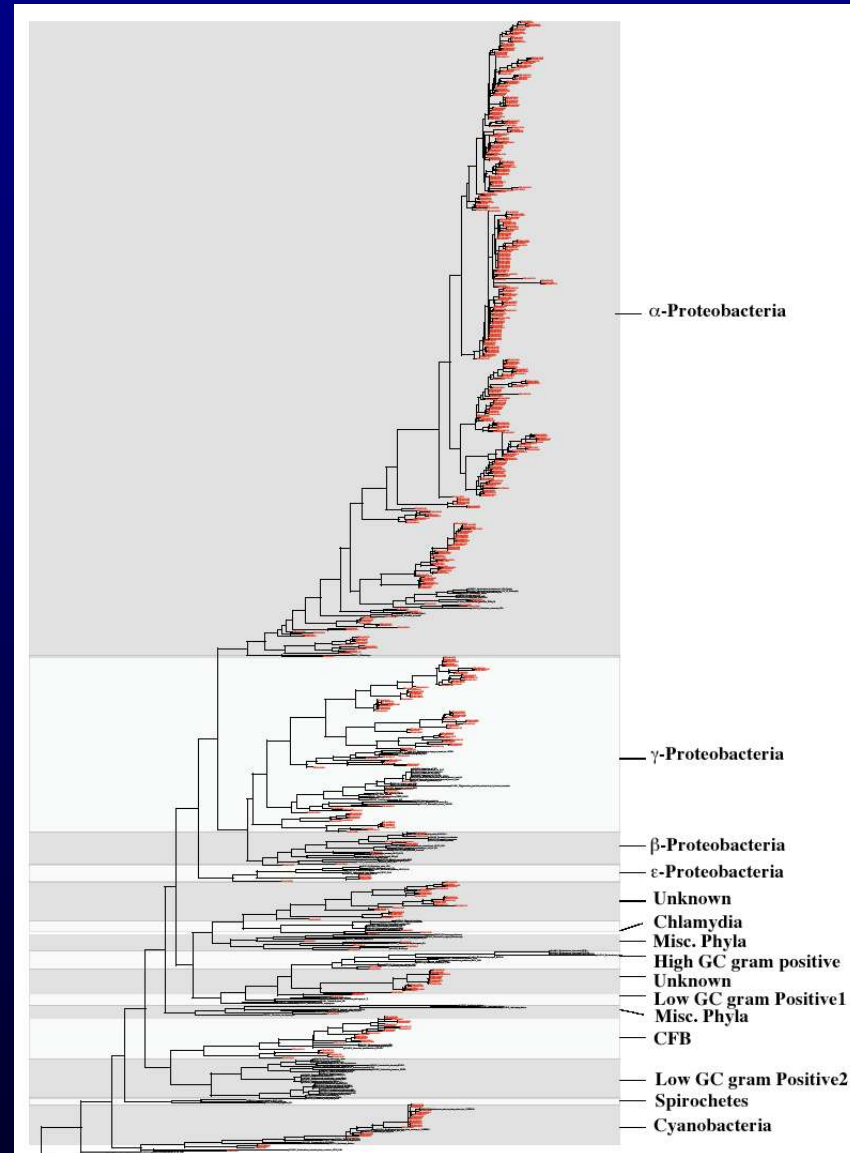
sequence



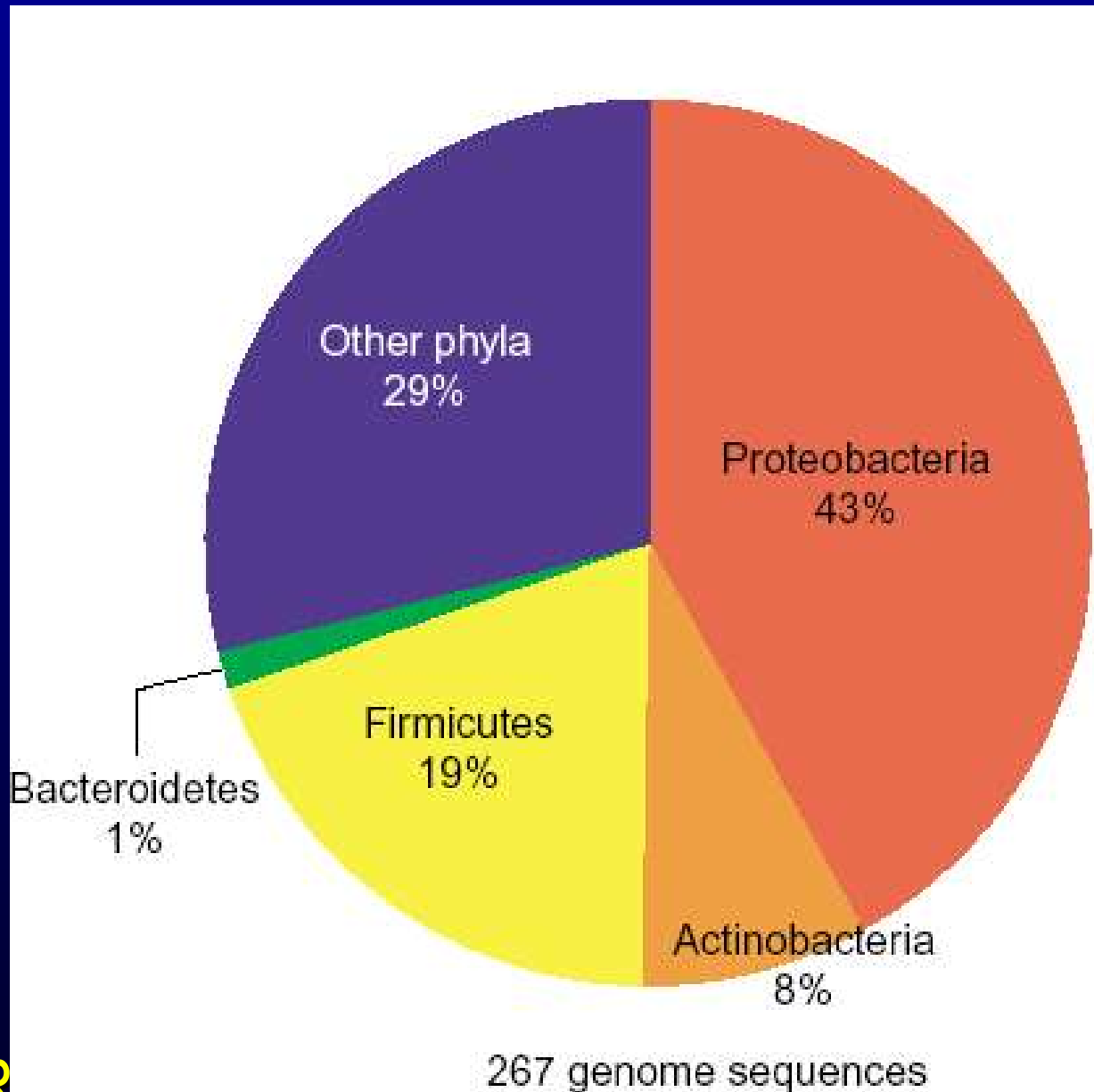
# rRNA as a Phylogenetic Anchor



# Shotgun Sequencing Allows Use of Alternative Anchors (e.g., RecA)



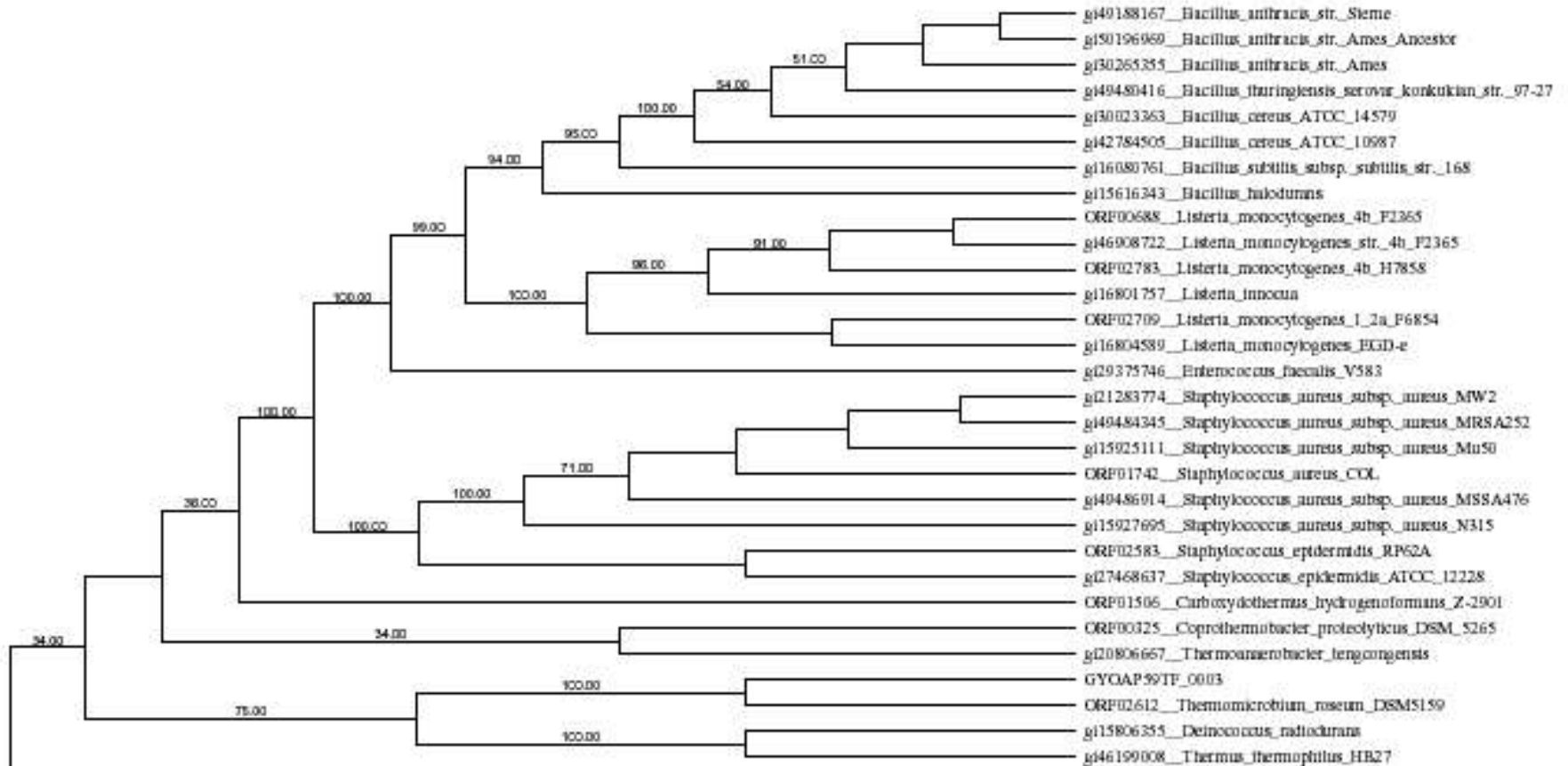
# Biased Sampling of Bacterial Genomes



Hugenholtz  
2002

Our Tree of Life  
Genomes Allow  
Anchoring of 100s  
of clones from  
Yellowstone Mats

# Tree of GYOAP59TF



# Blast of GYOAP59TF

>ORF02612-TG\_gtr\_857 rho transcription termination factor Rho  
    {Thermomicrobium\_roseum\_DSM5159}  
    Length = 426

Score = 1177 (419.4 bits), Expect = 2.9e-119, P = 2.9e-119  
Identities = 232/232 (100%), Positives = 232/232 (100%)

```
Query:      1  VAPIGRGQRGLIVSPPKAGKTVLLKHIANGITTNYKDIHLIVLLIGERPEEV TDMRRSVD 60
              VAPIGRGQRGLIVSPPKAGKTVLLKHIANGITTNYKDIHLIVLLIGERPEEV TDMRRSVD
Sbjct:    191 VAPIGRGQRGLIVSPPKAGKTVLLKHIANGITTNYKDIHLIVLLIGERPEEV TDMRRSVD 250

Query:      61  GEVISSTFDEPVEDHIRVAEMTLERAKRLVECGMDVVILLDSITRLARAYNLSVPPSGRT 120
              GEVISSTFDEPVEDHIRVAEMTLERAKRLVECGMDVVILLDSITRLARAYNLSVPPSGRT
Sbjct:    251 GEVISSTFDEPVEDHIRVAEMTLERAKRLVECGMDVVILLDSITRLARAYNLSVPPSGRT 310

Query:     121  LSGGIDPVALYPPKRFFGAARNIEGGGSLTIIATCLVDTGSRMDDVIYEEFKGTGNMELH 180
              LSGGIDPVALYPPKRFFGAARNIEGGGSLTIIATCLVDTGSRMDDVIYEEFKGTGNMELH
Sbjct:    311 LSGGIDPVALYPPKRFFGAARNIEGGGSLTIIATCLVDTGSRMDDVIYEEFKGTGNMELH 370

Query:     181  LDRKLAERRIFPAIDIQRSGTRREELLLDEQTLRQVWTMRRMVSM LGGTDGT 232
              LDRKLAERRIFPAIDIQRSGTRREELLLDEQTLRQVWTMRRMVSM LGGTDGT
Sbjct:    371 LDRKLAERRIFPAIDIQRSGTRREELLLDEQTLRQVWTMRRMVSM LGGTDGT 422
```



# Broader Impact

- Genome sequence data released to TIGR web site
- Genome Users Workshop to be held at TIGR in 2005
- Project used as model for evolution teaching in talks to HHMI, MD Science Teachers Association, and Montgomery County high school science teachers
- Project promoted to public through radio and print media

# Training

- Student involvement
  - Kevin Penn, technician at TIGR
  - Temylope Adeyefa-Olasupo, U. Md Undergrad working at COMB
  - Rebecca Brocato, high school intern, COMB
  - Joseph Wister, TIGR Summer Fellow with Karen Nelson
  - Ryan Corces-Zimmerman, high school intern working at TIGR
- Teaching using this project
  - MBL Molecular Evolution Workshop
  - MBL Genomics Workshop

# Future Issues

- New estimates suggests  $> 100$  bacterial phyla
- Someone needs to cover Archaeal diversity
- Need DNA repository